A Monte Carlo method for chemical potential determination in single and multiple occupancy crystals

Nigel B. Wilding[1] and Peter Sollich[2]

[1] *Department of Physics, University of Bath, Bath BA2 7AY, United Kingdom.*
[2] *King's College London, Department of Mathematics, Strand, London WC2R 2LS, United Kingdom.*

PACS 05.10.Ln –
PACS 61.50.Ks –
PACS 64.75.Xc –

Abstract – We describe a Monte Carlo scheme which, in a single simulation, yields a measurement of the chemical potential of a crystalline solid. Within the isobaric ensemble, this immediately provides an estimate of the system free energy, with statistical uncertainties that are determined precisely and transparently. An extension to multiple occupancy ("cluster") solids permits the direct determination of the cluster chemical potential and hence the equilibrium conditions. We apply the method to a model exhibiting cluster crystalline phases, where we find evidence for an infinite cascade of critical points terminating coexistence between crystals of differing site occupancies.

The phase behaviour of crystalline materials is important in fields as diverse as solid state physics, soft matter, mineralogy and pharmacology. For instance, metals and their alloys exhibit a rich variety of crystalline structures [1], including novel behaviour such as isostructural transitions [2]. Colloids can self-assemble into a variety of complex structures, some with applications in photonics [3]. Many drug compounds exhibit crystalline polymorphism which can influence their clinical function [4].

The staple simulation approach for predicting crystalline phase behaviour is via free energy estimates obtained by numerical integration along some path that connects the macrostate of interest to a reference state of known free energy [5, 6]. Such "Thermodynamic Integration" (TI) is popular because it is both conceptually simple and can be implemented with only a modest extension of the simulation framework needed for standard Monte Carlo (MC) sampling. However there are a number of respects in which it is less than ideal. The method hinges on the identification of a good path and reference macrostate. A 'good' path is short; but the reference macrostate (the choice of which is limited) may lie far from the physical macrostate of interest, entailing a large number of independent simulations to make the necessary link. A potentially more serious constraint on the path is that the derivative being measured should vary slowly, smoothly and reversibly along it; if it does not the numerical quadrature may be compromised. A phase transition en route is thus a particular hazard. Evidently one has to decide how many simulations are to be performed along the path and where. In so doing one must strike a suitable balance between minimizing computation time and ensuring that no region of the path is neglected. This may necessitate a degree of trial and error. The uncertainties to be attached to TI estimates are also problematic. Use of simple numerical quadrature will result in errors. Error bounds have to aggregate the uncertainties (statistical and systematic) associated with different stages of the integration process.

Beyond simple crystals, there is considerable interest in particles that self-assemble via microphase separation into periodically modulated nanostructures. Classic examples include the lamellar and micellar crystalline phases encountered in surfactants and copolymers [7], and the smectic phases of liquid crystals [8]. More recently, it has been discovered that when certain types of repulsive particles that lack a hard repulsive core are compressed to high density, multiple occupancy ("cluster") crystals are formed [9–11, 13]. Such coreless potentials serve as models for a wide range of soft matter systems such as star polymers, dendrimers and microgels in which particles can substantially overlap [14]. To describe cluster crystals one must allow for a crystalline lattice in which each lattice site can be occupied by multiple particles. Let us suppose that such a crystal has N_c lattice sites, labeled $i = 1...N_c$

and that site i is occupied by $n_c(i)$ particles (a "cluster"). Clusters are generally bi- or poly-disperse, so the total particle number is $N = \sum_{i=1}^{N_c} n_c(i) = N_c n_c$, with n_c the average occupancy. A fundamental problem for simulation is to determine the equilibrium values of n_c, the lattice parameter a, the pressure P and the chemical potential μ that correspond to some particle number density $\rho = N/V$ and temperature T of interest. As shown previously, measurement of the Helmholtz free energy F at fixed N_c in the constant-(NVT) ensemble is insufficient in this regard [10]. Instead one has to estimate the lattice site or cluster chemical potential μ_c, given by $N_c\mu_c = F + PV - \mu N$, which vanishes at equilibrium [15]. Doing so entails supplementing TI measurements of F with additional sampling of the chemical potential μ (via the Widom insertion method) and the pressure P (via the virial) [10]. This process, or alternatively a direct estimation of the constrained free energy [11], then has to be repeated for a range of values of n_c in order to pinpoint equilibrium. Accordingly it is cumbersome and laborious.

Here we introduce a new MC simulation scheme that allows direct calculation of the chemical potential of crystals (and thence the free energy) from a single simulation. Extending the method to cluster crystals permits direct estimation of the cluster chemical potential, while histogram reweighting techniques can be used to identify the equilibrium state without further simulation. We first describe the basic scheme for a simple crystal before outlining its cluster solid generalization.

The central idea is to construct (within the constant-NPT ensemble) a reversible sampling path between a lattice with $N + M$ particles and another with N particles. The relative probability of finding the simulation in the two states provides a measure of the Gibbs free energy difference $\Delta G = \mu M$. This yields the chemical potential μ, from which the Helmholtz free energy density follows immediately as $f = \mu\rho - P$, with P the prescribed pressure and ρ the measured particle number density.

To elaborate, consider the situation shown schematically in Fig. 1 for a cubic lattice (though note the method is applicable to any Bravais lattice). A constant-NPT ensemble Monte Carlo simulation [6] is to be found in one of two states $\alpha \in \{0, 1\}$. For $\alpha = 0$ the system comprises a periodic box of volume $V^{(0)}$ containing $(m + 1) \times m \times m$ unit cells of lattice parameter a. Each particle (a circle) is associated with a unique site of a fixed perfect lattice (a black dot): there are N particles in the cubic subvolume of m^3 unit cells shown, and $M = N/m$ particles in the remaining (rightmost) plane of unit cells. In the spirit of the phase switch method [16,17], we write the position vector of each particle i in terms of the displacement $\vec{u}_i$ from its lattice site $\vec{R}_i^{(0)}$, i.e. $\vec{r}_i^{(0)} = \vec{R}_i^{(0)} + \vec{u}_i$, $i = 1 \ldots N + M$.

The switch to the $\alpha = 1$ state comprises a reversible mapping in which the instantaneous particle displacements $\{\vec{u}_i\}$ are re-associated with a second set of lattice sites $\{\vec{R}^{(1)}\}$, such that $\vec{r}_i^{(1)} = \vec{R}_i^{(1)} + \vec{u}_i$. We set $\vec{R}_i^{(0)} = \vec{R}_i^{(1)}$ for $i = 1 \ldots N$. Thus for $\alpha = 1$, the first N lattice sites form a periodic cubic system with volume $V^{(1)} = V^{(0)} m/(m + 1)$, and the particles associated with these sites retain their relative positions within the box under the switch. By contrast, for the remaining M particles, the change in environment under the switch is more radical; they leave the box altogether to become "ghost" particles, associated with fixed sites $\{\vec{R}_i^{(1)}\}$, $i = N + 1 \ldots N + M$. Ghost particles are *independent* (so the relative positions of the fixed sites is arbitrary) and they experience only a harmonic confining potential $\phi_g(\vec{u}_i)$ whose amplitude is chosen to roughly match the average ghost particle displacement to that of real particles.

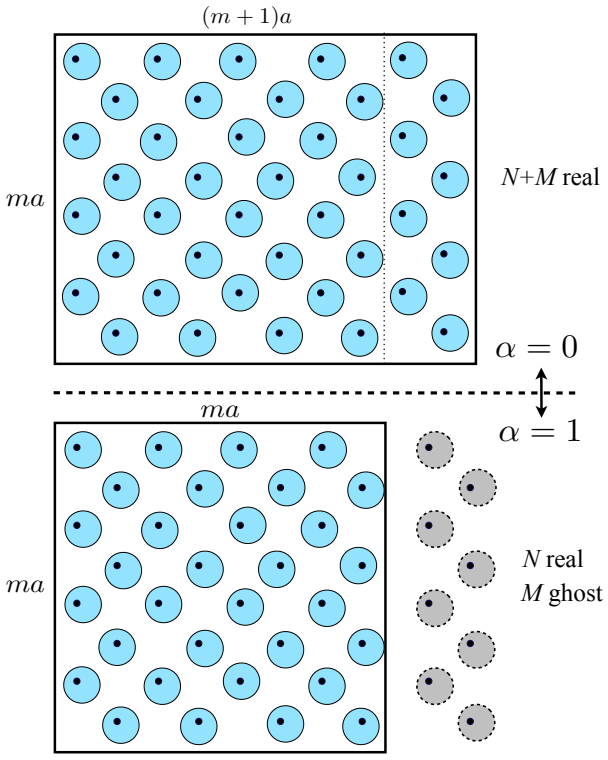


Fig. 1: Schematic of the ghost particle switching scheme described in the text.

Consider now the associated statistical mechanics. If we write the partition function of the system in state α as $Z^{(\alpha)}$, then since ghost particles are independent,

$$Z^{(0)} = Z(N + M, P, T)\,, \quad (1)$$

$$Z^{(1)} = Z_g^M Z(N, P, T)\,, \quad (2)$$

where $Z_g = \int d\vec{u} \exp(-\beta\phi_g(\vec{u}))$ is the (exactly calculable) partition function of one ghost particle, with $\beta = 1/(k_B T)$ and k_B Boltzmann's constant which we henceforth set equal to unity. The free energy change associated with the switch $\alpha = 1 \to 0$ follows as

$$\beta\Delta G = \beta M\mu = \ln \frac{Z(N, P, T)}{Z(N + M, P, T)} = \ln \frac{Z^{(1)}}{Z^{(0)} Z_g^M}\,, \quad (3)$$

with μ the chemical potential. In order to estimate ΔG, we supplement standard MC updates of the particle displacements and box volume with attempts to switch α. These are accepted with the standard (NPT) ensemble probability: $p_a = \min[1, \nu \exp(-\beta(P\Delta V + \Delta E))]$, with $\nu = V^{(1-\alpha)}/V^{(\alpha)}$. In general, however, such switch attempts suffer low acceptance rates since the particle displacements $\{\vec{u}_i\}$ for the current α may not all be typical for the switched α. To deal with this we implement biased ("umbrella") sampling [6], which enhances the probability of configurations (for each α) for which the instantaneous switch cost $\kappa = \beta(p\Delta V + \Delta E)$ is small. Specifically we include a weight function $\eta_\alpha(\kappa)$ in the acceptance probabilities for all MC updates. Weights are obtainable via any of the standard techniques such as transition matrix or Wang Landau sampling. Their effects are unfolded from the sampling in the usual way [6] at the end of the simulation.

In this manner one accumulates the relative probability $p^{(1)}/p^{(0)} = Z^{(1)}/Z^{(0)}$ of finding the simulation in the respective α states, from which the requisite chemical potential follows as

$$\mu = \beta^{-1}[M^{-1}\ln(p^{(1)}/p^{(0)}) - \ln Z_g]\,. \tag{4}$$

Statistical errors in μ are determined simply by the switching statistics and readily quantified via a block analysis. If the uncertainty in $p^{(1)}/p^{(0)}$ is δ, then that in μ is $O(\delta/M)$. Since M is typically $O(10^2)$, this bestows the method with high sensitivity. To test it we have measured μ for a Lennard-Jones fcc crystal of system size $m = 4$. Interactions were truncated at $r_c = 2.9$ and a mean-field correction of the usual type [6] was applied. As no previous estimates of chemical potentials for this system exist (to our knowledge), we use our results to calculate the absolute Helmholtz free energy density $f = \mu\rho - P$, and compare with literature estimates determined by TI. Only one such previous estimate quotes an associated uncertainty, allowing for meaningful comparison [18]. For the state point $T = 2.0, \rho = 1.28$, we find $\beta P = 20.985(2), \beta\mu = 18.967(3), \beta f = 3.292(1)$, the latter comparing well with the estimate of Vega and Noya: $\beta f = 3.290(4)$, though our error bar is substantially smaller. In all, our measurement consumed circa 24 hours of CPU time, comprising 5×10^6 displacement attempts per particle and 10^7 volume change attempts.

We next address a challenging and fundamental problem in the simulation of cluster solids. At some (T, ρ), equilibrium corresponds to a particular value of the cluster occupancy n_c and the lattice parameter a. But specifying ρ fixes neither of these parameters, e.g. in an fcc cluster crystal, $\rho = 4n_c/a^3$, which can be realized by many combinations of n_c and a. In a real system, N_c changes to relax the system to equilibrium. However, in conventional simulation ensembles, the value of N_c is constrained on accessible times scales by free energy barriers and does not fluctuate. As described above, one approach to determining equilibrium in these circumstances is to estimate the cluster chemical potential $\mu_c(n_c)$ via a laborious combination of TI, Widom particle insertion and virial sampling [10], while another is to directly estimate the constrained free energy via TI [11]. A third approach uses TI in combination with a so-called [N]pT ensemble which allows for histogram reweighting in the volume and particle number [12]. However, all of these methods rely on thermodynamic integration and therefore potentially suffer from the drawbacks mentioned above. By contrast ghost particle switching (or more precisely ghost *cluster* switching) provides a simpler, more elegant and potentially more efficient solution to this problem.

In seeking to apply the method, it is expedient both in terms of capitalizing on histogram reweighting and for directly estimating the cluster chemical potential, to employ an ensemble in which both n_c and a are free to fluctuate– the constant μ, P, T ensemble. This ensemble is rarely utilized in simulations because the extensive scaling of the entropy means that the partition function is finite for high pressures, diverges on approach to equilibrium, e.g. $Z(\mu, P, T) \sim (P - P_{\rm eq})^{-1}$ as $P \to P_{\rm eq}^+$, and is infinite for all lower pressures. This is no longer a problem when we have a constraint of fixed N_c. The partition function is then

$$Z(\mu, P, T, N_c) = \sum_N \int dV\, dE\, e^{S(N,V,E,N_c) - \beta(E+PV-\mu N)} \tag{5}$$

The entropy will be extensive for large N_c, $S(N, V, E, N_c) = N_c s(n_c, v_c, e_c)$ with $n_c = N/N_c$, $v_c = V/N_c$, $e_c = E/N_c$. The dominant contribution to Z comes from the maximum of the integrand, where $-\beta\mu = \partial s/\partial n_c$, $\beta P = \partial s/\partial v_c$ and $\beta = \partial s/\partial e_c$. Denoting these saddle point values with an asterisk, the extensive contribution to $\ln Z$ is

$$\ln Z(\mu, P, T, N_c) = N_c[s(n_c^*, e_c^*, v_c^*) - \beta(e_c^* - \mu n_c^* + P v_c^*)] \tag{6}$$

In general, this is a non-equilibrium partition function because at equilibrium any two of (μ, P, T) determine the third. To find the equilibrium condition, assume (μ, P, T) is a set of equilibrium parameters, then so must n_c^*, v_c^* and e_c^* be. This means that they are obtained by maximizing the entropy $S(N, V, E, N_c)$ over N_c. From the extensive form of S above, and using that $\beta = \partial s/\partial e_c$ etc, this shows directly that the combination in square brackets in (6) must vanish.

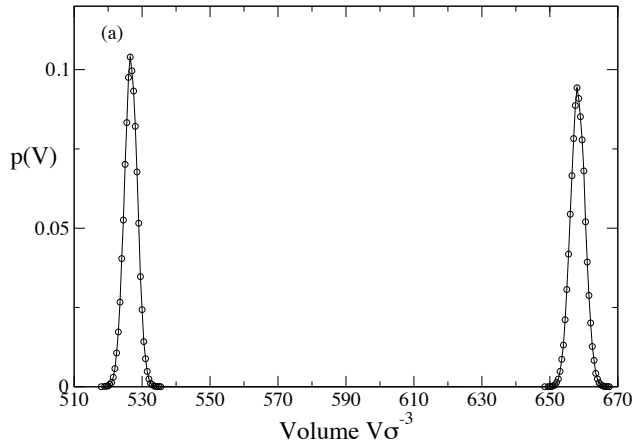
The upshot of this analysis is that $-\beta^{-1}\ln Z(\mu, P, T, N_c)/N_c$ vanishes at equilibrium, which is as expected given that by standard thermodynamic arguments this quantity can also be identified with the cluster chemical potential μ_c. On the other hand, $Z(\mu, P, T, N_c)$ is also the weight of different N_c values in a simulation where N_c can fluctuate, and so the above equilibrium criterion tells us that at equilibrium these weights are to leading order independent of N_c.

Ghost cluster switching allows one to repeatedly add and remove a crystallographic plane of lattice sites [1], thereby circumventing the barriers between different values of N_c. Measurements of the relative probabilities of macrostates with different N_c then directly probes μ_c.

The implementation is technically similar to that described for simple crystals, except that the particle number is permitted to fluctuate via insertions/deletions which for cluster solids are efficient owing to the lack of a repulsive hard core in the potential. For ghost sites, in addition to choosing the harmonic amplitude such that ghost particle displacements are similar to those of real particles, we impose a ghost chemical potential μ_g chosen to yield an average site occupancy close to that of real sites. Monte Carlo moves include sweeps over lattice sites in which attempt particle displacements, particle transfers. Volume dilations are also attempted. The order parameter against which we bias to enhance the switch probability is extended from the single occupancy case to become $\kappa = \beta(P\Delta V + \Delta E \pm (\mu - \mu_g)N_g)$, with N_g the instantaneous number of particles associated with lattice sites $i = N+1 \ldots N+M$. We sample the distribution of the observables (N, V, E) across the two values of α, from which we unfold the effects of the biasing weights and the ghost particle free energy. This yields (*inter alia*), the volume distribution $p(V|\mu, P, T)$, which exhibits two peaks, one corresponding to $\alpha = 1$ and the other for $\alpha = 0$. Equilibrium is signaled in a very simple fashion by the equality of the peak areas of $p(V)$, as these have the same ratio as $Z(\mu, P, T, N_c)$ and $Z(\mu, P, T, N_c + M_c)$.

To validate the methodology we have considered a prototype cluster solid: the generalized exponential model (GEM-4) whose interaction potential is $u(r) = \epsilon \exp((-r/\sigma)^4)$. We determined $P_{\rm eq}$ and $\mu_{\rm eq}$ for $T = 1.1, \rho = 8.5$ in the fcc phase – a state point for which prior TI data is available [10]. The initial simulation was performed for $P = 114.5$, $\mu = 29.7$ for a system size of $m = 4$ and the resulting probability distribution $p(V)$ was subsequently extrapolated in P and μ to yield both equal peak areas and a density matching the target value. Fig. 2 shows the resulting equilibrium form, obtained for $P_{\rm eq} = 114.46(1)$, $\mu_{\rm eq} = 29.753(3)$. The associated cluster number is $n_c = 17.470(5)$ and the fcc lattice parameter $a/\sigma = 2.018(1)$. We note that these results agree to within error with those of Mladek [19], although they were obtained with a computational expenditure of circa 100 CPU hours, $1-2$ orders of magnitude faster than the original method. More precisely, our results were obtained for circa 10^6 displacement attempts per particle, 10^6 transfer attempts per lattice site, and 5×10^5 volume change attempts. Fig. 2 shows a portion of a snapshot of an equilibrium configuration colored by cluster.

[1] Note that since N fluctuates in additions to N_c, our method achieves the maximum resolution of $1/N_c$ in its estimates of the cluster occupancy $n_c = N/N_c$; Zhang and Charbonneau [12] express the misconception that our approach is constrained in this respect and suffers accuracy problems.



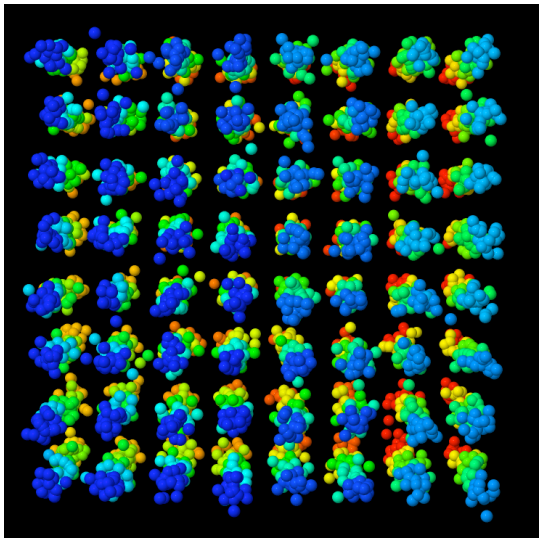


Fig. 2: **(a)** Equality of peak areas in $p(V)$ that signifies equilibrium in the (μ, P, T) ensemble (see text for the equilibrium parameters). (b) An equilibrium configuration viewed along the [100] direction.

As an application we consider a recent proposal [13] concerning the existence of a cascade of low temperature critical points in the GEM-4 model. On the basis of ground state energy calculations and a phonon analysis, these authors found a sequence of low temperature isostructural (fcc) phase transitions on increasing density. At $T = 0$, each transition is such that the cluster number changes by unity from one integer value to the next, i.e. $n_c \to n_c + 1$. No theoretical evidence was found that any of these transition has an associated critical point at finite temperature, but the authors hypothesized that this should be the case. Indeed subsequent evidence for a critical point terminating the lowest density transition $n_c = 2 \leftrightarrow 3$ has been found using TI [11,12].

We have used the present method to search for further critical points in the model at higher density. We employed the known universal Ising form of the critical order parameter distribution to estimate the first four critical points [20] using a system size $m = 3$. The resulting

critical point parameters[2] are listed in table 1. Surprisingly we find that T^c is equal within error in each case as reflected in the independence of the form of the density distributions at $T = 0.04348$ shown in Fig. 3. Although this finding appears to be linked to the n_c independence of the density difference between the coexisting phases at $T = 0$ that is apparent in ref. [13], a deeper understanding will require fresh insight into the interplay between fluctuations in site occupancy and unit cell dimensions in cluster crystals.

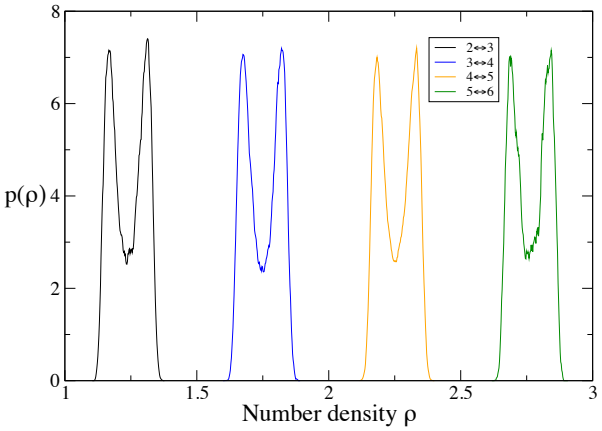


Fig. 3: Density distributions of the GEM-4 model corresponding to the near-critical point parameters listed in table 1.

In summary we have introduced an efficient and accurate 'ghost particle switching' method for chemical potential determination in crystalline solids within the constant-NPT ensemble. The method, which circumvents the need for integration to distant references states and its attendant pitfalls, requires only a single simulation at the state point of interest and yields statistical uncertainties directly and transparently. Such access to the chemical potential permits the direct determination of phase boundaries by matching of μ and P in the coexisting phases. An extension of the method to multiple occupancy crystals within the constant-μPT ensemble simplifies the problem of determining their equilibrium parameters. As a demonstration of its power in this regard, we have studied the GEM-4 cluster solid, uncovering the presence of a cascade of iso-temperature critical points. More generally, the basic approach of ghost particle switching should be applicable to any system exhibiting periodic micro- and mesophase separation in which the repeat unit can contain many individual particles, examples being smectic liquid crystals and the lamellar and crystalline micellar phases that appear in surfactants, copolymers and colloids with competing attractive and repulsive interactions.

[2]The apparent T^c that we find for the $n_c = 2 \leftrightarrow 3$ transition is some 5% lower for our system size ($m = 3$, $N \approx 270$) than that reported in Ref. [12] for a system of $N = 500$ particles. However this discrepancy is consistent with the scale and character of finite-size effects reported in [12].

* * *

We thank Bianca Mladek and Rob Jack for fruitful discussions.

REFERENCES

[1] D.A. Young, *Phase Diagram of the Elements* (University of California press, Berkeley, 1991).
[2] F. Decremps *et al.*, Phys. Rev. Lett. **106**, 065701 (2011).
[3] Y.-H. Ye, F. LeBlanc, A. Hache, and V.-V. Truong, Appl. Phys. Lett. **78**, 52 (2001).
[4] *Polymorphism in Pharmaceutical Solids*, edited by H.G. Brittain (Marcek Dekker, New York, 1999).
[5] D. Frenkel and A.J.C Ladd, J. Chem. Phys. **81**, 3188 (1984).
[6] D. Frenkel and B. Smit, *Understanding Molecular Simulation* (Academic, San Diego, 2002).
[7] *Amphiphilic block copolymers: Self-assembly and applications*, edited by P. Alexandridis and B. Lindman (Elsevier, Amsterdam, 2000).
[8] P.J. Collings, *Liquid Crystals: Nature's Delicate Phase of Matter* (Princeton University Press, Princeton, 2001).
[9] B. M. Mladek *et al.*, Phys. Rev. Lett. **96**, 045701 (2006).
[10] B. M. Mladek, P. Charbonneau, and D. Frenkel, Phys. Rev. Lett. **99**, 235702 (2007).
[11] K. Zhang, P. Charbonneau, and B. M. Mladek, Phys. Rev. Lett. **105**, 245701 (2010).
[12] K. Zhang and P. Charbonneau, Phys. Rev. E **86**, 042501 (2012).
[13] T. Neuhaus and C. N Likos, J. Phys.: Condens. Matter **23**, 234112 (2011).
[14] C.N. Likos, Soft Matter **2**, 478 (2006).
[15] W. C. Swope and H. C. Andersen, Phys. Rev. A **46**, 4539 (1992).
[16] A. D. Bruce, N. B. Wilding, and G. J. Ackland, Phys. Rev. Lett. **79**, 3002 (1997).
[17] N. B. Wilding and A. D. Bruce, Phys. Rev. Lett. **85**, 5138 (2000).
[18] C. Vega and E. G Noya, J. Chem. Phys. **127**, 154113 (2007).
[19] B.M. Mladek, Ph.D. thesis, T.U. Wien, 2007.
[20] N. B. Wilding, Phys. Rev. E. **52**, 602 (1995).

Transition	$\rho^{\rm c}$	$n_{\rm c}^{\rm c}$	$P^{\rm c}$	$\mu^{\rm c}$	$T^{\rm c}$
$2 \leftrightarrow 3$	1.239(4)	2.495(5)	1.974(2)	2.9151(5)	0.0435(4)
$3 \leftrightarrow 4$	1.740(4)	3.490(6)	3.879(2)	4.1878(4)	0.0435(4)
$4 \leftrightarrow 5$	2.257(5)	4.495(6)	6.418(2)	5.4521(5)	0.0435(4)
$5 \leftrightarrow 6$	2.762(5)	5.486(7)	9.575(3)	6.709(5)	0.0435(5)

Table 1: Estimated values of the apparent critical parameters based on matching to the universal finite-size order parameter distribution [20] for a system size $m = 3$. Data is given for the critical density $\rho^{\rm c}$, lattice site occupancy $n_{\rm c}^{\rm c}$, pressure $P^{\rm c}$, chemical potential $\mu^{\rm c}$ and temperature $T^{\rm c}$, for the first four members of the infinite cascade of phase transitions having $n_{\rm c} \leftrightarrow n_{\rm c} + 1$ at $T = 0$.